Revision of Diagnostic Logic Using a Clinical Database

PETER HAUG, MD, PAUL D. CLAYTON, PhD, PAMELA SHELTON, MS, TRACY RICH, IRENA TOCINO, MD, PHILIP R. FREDERICK, MD, ROBERT O. CRAPO, MD, WILLIAM J. MORRISON, MD, HOMER R. WARNER, MD, PhD

Statistical pattern-recognition techniques have been frequently applied to the problem of medical diagnosis. Sequential Bayesian approaches are appealing because of the possibility of generating the underlying sensitivities, specificities, and prevalence statistics from the estimates of medical experts. The accuracy of these estimates and the consequences of inaccuracies carry implications for the future development of this type of system. In an effort to explore these subjects, the authors used statistics derived from a clinical database to revise the diagnostic logic in a Bayesian system for generating a differential diagnostic list. Substantial changes in estimated a priori probabilities, sensitivities, and specificities were made to correct for significant under- and overestimations of these values by a group of medical experts. The system based on the derived values appears to perform better than the original system. It is concluded that the statistics used in a Bayesian diagnostic system should be derived from a database representative of the patient population for which the system is designed. *Key words:* diagnosis; computer-assisted; Bayes theorem; lung diseases. **(Med Decis Making 1989;9:84–90)**

Many medical diagnostic systems have been developed over the past 30 years. A substantial number of these systems rely on statistical algorithms to estimate the respective likelihoods of a group of diseases. While a variety of techniques employing discriminant functions have been tested,[1,2] statistical systems based on Bayesian probability analysis have been and continue to be popular.[3-6]

Bayesian systems have an advantage over discriminant functions in that the parameters required for their function are disease prevalences (a priori probabilities) and the familiar sensitivity and specificity that describe the association between clinical data and the disease categories represented in the system. These concepts are known to physicians, and they will often be comfortable in estimating the values of these parameters based on their personal experience. In addition, some information bearing on sensitivity and specificity can be found in the medical literature. Thus, it is possible to develop a medical diagnostic program using values provided by one or more medical experts, supplemented by estimates from the medical literature.

Received August 21, 1988, from the Departments of Medical Informatics, Radiology, and Pulmonary Medicine, LDS Hospital and the University of Utah, Salt Lake City, Utah. Revision accepted for publication December 8, 1988. Presented in part at the Sixth Annual Joint National Congress of AAMSI (1987), San Francisco. Supported in part by Grant 2-R18-HS03810/04, National Center for Health Services Research.

Address correspondence and reprint requests to Dr. Haug: Department of Medical Informatics, LDS Hospital, 8th Avenue and C Street, Salt Lake City, UT 84143.

Unfortunately, there are pitfalls in this approach. Tversky and Kahneman[7] have pointed out the general human difficulty with estimating probabilities. While physician estimates appear accurate in some studies,[8,9] there is considerable evidence that they differ from reality in many situations.[10,11] In addition, both physician estimates and values found in the medical literature can suffer from differences in the underlying prevalences and presentations of disease in different patient populations. This affects the portability of systems based on the estimates of physicians from one location to other settings.

Another source of statistics useful for developing Bayesian diagnostic systems exists. It is a clinical database containing patient information useful in the

Table 1 • Diseases for Which Diagnostic Frames Were Initially Constructed

Acute bronchitis	Histiocytosis X
Asbestosis	Hodgkin's disease
Aspiration pneumonia	Influenza
Asthma	Lung abscess
Bacterial pneumonia	Metastatic neoplasm
Bronchiectasis	Non-Hodgkin's lymphoma
Chronic bronchitis	Primary pulmonary neoplasm
Coal-worker's pneumoconiosis	Primary pulmonary hypertension
Coccidioidomycosis	Pulmonary embolism
Congestive heart failure	Sarcoidosis
Diffuse idiopathic fibrosis	Silicosis
Drug-related pneumonitis	Spontaneous pneumothorax
Emphysema	Tuberculosis
Goodpasture's syndrome	Wegener's granulomatosis
No pulmonary disease	

```
(1) FRAME 1    =.== PNEUMONIA (HISTORY)

    FINAL EVALUATION:
(2) A VAL: M

    SECTOR LOGIC:
(3) A ARITH:    0.014

(4) B SEARCH: ^ (A) HAVE YOU HAD RECENT CHEST PAIN?

    C SEARCH:   (A) HAVE YOU HAD A FEVER WITH THIS ILLNESS?

    D SEARCH:   (A) HAVE YOU HAD CHILLS WITH THIS ILLNESS?

    E SEARCH: ^ (A) HAVE YOU HAD A COUGH WITH THIS ILLNESS?
    F SEARCH:   (A) IS YOUR CHEST PAIN INCREASED BY BREATHING DEEPLY?
                (B) IS YOUR CHEST PAIN INCREASED BY COUGHING?
                USE ANSWER MAX(A, B)

    G SEARCH:   (A) HAVE YOU BEEN SHORT OF BREATH WITH THIS ILLNESS?

    H SEARCH:   (A) IS YOUR SPUTUM YELLOW, GREEN OR BROWN?

(5) I PROB:     A,  IF ex: C OR D,  USE val: MAX(C, D)
          ANSWER: (N, Y),  TRUE: (0.15, 0.85),  FALSE: (0.7, 0.3)

    J PROB:     I,  IF ex: E,  USE val: E,  ANSWER: (N, Y)
         TRUE:    (0.1, 0.9),  FALSE: (0.8, 0.2)

    K PROB:     J,  IF ex: F,  USE val: F,  ANSWER: (N, Y)
         TRUE:    (0.71, 0.29),  FALSE: (0.9, 0.1)

    L PROB:     K,  IF ex: G,  USE val: G,  ANSWER: (N, Y)
         TRUE:    (0.56, 0.44),  FALSE: (0.87, 0.13)

    M PROB:     L,  IF ex: H,  USE val: H,  ANSWER: (N, Y)
         TRUE:    (0.35, 0.65),  FALSE: (0.95, 0.05)
```

FIGURE 1. Parts of a diagnostic frame for the computer-directed history: (1) frame label, (2) final evaluation slot, (3) a priori probability for this disease, (4) data specification; indicates the questions required to calculate disease likelihood, (5) specification of statistics (sensitivity and specificity) associated with yes and no answers to referenced question.

diagnosis of a target group of diseases. DeDombal et al. have demonstrated the power of a system developed from this sort of information to accurately suggest diagnoses in patients with acute abdominal discomfort.[4] With the advent of large, clinically oriented medical information systems the opportunity for creating systems based on statistics extracted from accumulated clinical data is markedly increased.

We have used the medical decision support tools in the HELP hospital information system to develop a computerized representation of the diagnostic logic for generating a brief differential diagnostic list in the first day of hospitalization. This system uses a Bayesian approach to evaluate the likelihoods of a group of pulmonary diseases. The logic and statistics required to describe each disease are organized into modules referred to as diagnostic frames. The original statistical underpinnings of this system were a set of a priori probabilities, sensitivities, and specificities generated by a group of physicians. We describe the use of a clinical database to revise these statistics and the effect of this revision on the function of the diagnostic system.

Methods

The model of diagnosis used for the projects described above is based on a modified sequential Bayesian approach. The multi-membership version of Bayes equation is used[12] (see appendix). A system using this model may be sensitive to the accuracy of the statistics used in the frames. We determined to test the hypothesis that revising the estimates of our physician panel through the analysis of the clinical database would improve the accuracy of the diagnostic frames.

To build a knowledge base for these experiments, a group of five physicians was assembled. This group consisted of a specialist in pulmonary medicine, an internist, and three radiologists with special interest in chest radiology. They chose the set of diseases to include in the system and provided a list of the patient data most useful in the diagnosis of these diseases. They also assisted in assigning initial probability estimates for each of the manifestations. The statistical estimates produced by this group were supplemented with information from the medical literature and by

values derived from clinical data collected as part of an experimental history tested some years ago.[3]

Since a sequential, Bayesian approach to diagnosis was chosen, the knowledge we obtained from these physicians was in the form of probability estimates; an a priori probability for each disease in the inpatient population and a sensitivity and specificity for each manifestation in each disease. These estimates were supplemented by data from the medical literature, and by a review of pertinent information derived from the original, hard-coded, Bayesian history system.

A group of 28 diseases was modeled in this way (table 1). An additional diagnostic module was created to explicitly identify patients with no pulmonary disease. Figure 1 is an example of a module for the diagnosis of pneumonia. These frames formed the knowledge base used in both tests of approaches to collecting the patient history.

The statistics required for the functioning of the diagnostic frames are 1) a set of a priori probabilities representing the expected prevalences of individual diseases, 2) a sensitivity for each manifestation of each disease, and 3) a specificity for each manifestation of each disease. In the original disease frames these values were estimated by a group of physicians as described above. When relevant statistical information was readily available in the medical literature, these values were used to assist the physicians in their estimations. In addition, for those statistics representing the association between elements of the history and the diseases, they had access to values generated as a part of an earlier study.[3] However, these values consisted of only the sensitivities linking appropriate historical manifestations to a group of common pulmonary diseases.

Our goal was to use information stored in the HELP clinical database to revise these statistics. We concentrated on the historical and radiographic data since these were felt to be the greatest contributors to the diagnostic process for the pulmonary diseases with which we were concerned.

To evaluate the hypothesis, information was collected prospectively for a group of 627 patients entering the LDS Hospital. In order to assure an adequate number of pulmonary diseases for analysis, only patients who had a chest x-ray ordered in the first 48 hours were included. The information usually captured by the HELP system was supplemented with a patient history gathered with a paper questionnaire. The questionnaire was designed specifically to capture the information used in the diagnostic modules. "Yes" or "no" answers reflecting the presence or absence of 182 symptoms were required for the functioning of these modules.

A terminal-based, interactive questionnaire was developed to collect a descriptive report of each patient's initial chest x-ray. This information is more detailed than the usual radiology report captured for the HELP database by our standard radiology reporting system. All 627 of the patients had their initial chest x-rays entered by radiologists using this system.

This group of patients was then divided into a training set consisting of 527 patients and a test set consisting of 100 patients. The training set was used to revise the statistics in the diagnostic frames and the test set was reserved to evaluate changes in diagnostic accuracy associated with the use of the new versions of these frames.

Diagnoses were assigned to individual patients by examining the final discharge diagnoses stored in the clinical database. This determination was based upon the ICD-9[13] codes, which are selected by the patients' attending physicians at the time of discharge and are entered into the computer by the medical records department. They reflect the opinion of the clinician at a time when data from the entire hospitalization is available. If none of the 28 diseases for which we had developed decision logic were present, the patient was designated as having "no pulmonary disease." These were patients hospitalized for diseases in other organ systems.

Next, the 527-member training set was analyzed to provide sample-based estimates of the statistics used in the frames. Database analysis tools in the HELP system were used to extract patients with specified diseases and examine relevant clinical data. Where adequate numbers of patients had had a disease, an a priori probability and relevant sensitivities and specificities were generated. A minimum of six patients with a given disease was required before a new a priori probability and sensitivities were derived from the database. Since specificities are based on the set of patients without a disease, new specificities could be derived for the findings in all of the disease modules.

After analyzing the training set, we revised the a priori probabilities as well as the sensitivities and specificities for the history data and the radiology findings represented in the frames. The revision process consisted of replacing the original, estimated statistics with those derived from the training set wherever possible. Ten of the 29 diseases were represented by sufficient patients to allow a complete revision. For the remainder, only the specificities of the manifestation/disease combinations could be revised.

To examine the alterations in these values, we divided them into sensitivities and specificities associated with history and those associated with findings on chest x-ray. For the sensitivities, only the ten most common diseases were analyzed. All diseases were included in evaluating the specificities. We compared the estimated and revised statistics to determine the accuracy of our estimates and the comparative accuracies of the estimates for history and radiology. The

a priori probabilities were also compared with earlier estimates.

We chose to evaluate the significance of estimation errors by examining their effects on the accuracy of the diagnostic system. Following revision of the diagnostic logic, we compared the diagnostic results of the original frames with those of the revised frames. The system was tested in three ways. First, the diagnostic frames were limited to the history data available for each patient; second, the frames were allowed to access only the chest x-ray data for each patient; and third, the diagnostic system was tested with both the history and the x-ray data. In each case, both the frames based on the original statistical estimates and similar frames using the derived statistics were evaluated against the same data set. To accomplish this, all 29 frames were run against each patient in the 100-member test set, and a one-to-five-member differential diagnostic list was constructed. This list consisted of the most likely diagnoses but excluded any disease with a likelihood less than 1%.

The final discharge diagnoses of the patients were then examined and each computer-generated differential diagnostic list was compared with this group of diseases. The computer's differential list was considered accurate when it contained a known discharge diagnosis and inaccurate when a diagnosis was missed.

McNemar's test was used to compare the accuracy of the differential lists produced by the original frames with that of those produced using the revised frames. Paired t-tests were used to compare the original and derived statistics.

Table 2 • Comparison of Ten Original and Derived Prior Probabilities

Sector	Original a Priori Probability	Derived a Priori Probability
Pneumonia	0.014	0.067
Pulmonary embolus	0.012	0.022
Acute bacterial bronchitis	0.008	0.013
Aspiration pneumonia	0.004	0.020
Chronic bronchitis	0.014	0.054
Emphysema	0.014	0.025
Asthma	0.013	0.035
Pulmonary neoplasm	0.005	0.030
Pulmonary edema (CHF)	0.014	0.125
No pulmonary disease	0.500	0.650
Total ± SD	0.060 ± 0.155	0.104 ± 0.194*

*$p < 0.05$ (paired t-test).

Table 3. • Comparison of the Original and Derived Sensitivities and Specificities: Changes in the Conditional Probabilities for History and X-ray Results

	Original Statistics (Mean ± SD)	Derived Statistics (Mean ± SD)
History		
Sensitivity (n = 76)	0.456 ± 0.272	0.450 ± 0.239
Specificity (n = 202)	0.848 ± 0.134	0.802 ± 0.171†
X-ray		
Sensitivity (n = 20)	0.474 ± 0.264	0.367 ± 0.262*
Specificity (n = 78)	0.896 ± 0.129	0.941 ± 0.096†

*$p < 0.05$ (paired t-test).
†$p < 0.001$ (paired t-test).

Results

There were a total of 110 discharge diagnoses in these patients, of which 58 were pulmonary diseases. The analysis was done both with and without the 52 patients with no pulmonary disease. The differential diagnostic lists were evaluated to determine whether they contained each recorded discharge diagnosis (if none of the other diseases were present as discharge diagnoses then "no pulmonary disease" was considered appropriate). In addition, the diagnostic lists were examined to determine whether the actual discharge diagnosis appeared as the leading alternative on each list.

Table 2 and 3 compare the a priori probabilities and the sensitivities and specificities for relevant groups of clinical findings used in the original and revised frames. Table 2 gives the original and derived prior probabilities for the ten most common diseases. Values are rounded to three decimal places. Some patients had more than one disease, resulting in a mean for the derived values that was greater than 0.10.

Shown in table 3 are the means and SDs for the original and derived sensitivities and specificities. We have divided the figure into statistics involving history data and those involving chest x-ray data. Analyses of changes in these conditional probabilities were based on 76 original and derived historical sensitivities, 202 original and derived historical specificities, 20 original and derived radiologic sensitivities, and 78 original and derived radiologic specificities. The difference in the numbers of sensitivities and specificities analyzed reflects the fact that only ten diseases were represented by numbers of patients adequate for the analysis of sensitivities; specificities could be analyzed for each finding in each disease. All but one of the differences noted between the means for the measured and estimated statistics were significant at the 0.05 level.

Table 4 gives representative examples of the changes in statistics for two of the diseases. Congestive heart failure and pneumonia were the two most frequent illnesses (excepting no pulmonary disease) represented in the 527-member training population. The original and derived a priori probabilities, sensitivities, and specificities are indicated.

For the ten prior probabilities shown we found evidence of significant underestimate by the experts. In

Table 4 • Example of Revisions in the Statistics for Congestive Heart Failure and Pneumonia

	Original	Derived	Original Sensitivity	Derived Sensitivity	Original Specificity	Derived Specificity
Congestive heart failure						
A priori probability	0.014	0.125				
History						
Exertional dyspnea			0.70	0.64	0.80	0.55
Orthopnea			0.55	0.36	0.92	0.84
Paroxysmal nocturnal dyspnea			0.43	0.33	0.90	0.87
Pedal edema			0.60	0.66	0.90	0.68
Nocturia			0.47	0.51	0.70	0.62
Cardiac medications			0.70	0.52	0.70	0.73
Digoxin			0.60	0.49	0.80	0.90
Heart murmur			0.37	0.39	0.81	0.84
Prior heart failure			0.40	0.43	0.90	0.95
X-ray						
Cardio-pericardial enlargement or pulmonary venous hypertension			0.80	0.80	0.90	0.85
Diffuse pulmonary infiltrates			0.60	0.39	0.80	0.91
Perihilar infiltrates			0.60	0.47	0.95	0.99
Pneumonia						
A priori probability	0.014	0.067				
History						
Exertional dyspnea			0.70	0.64	0.80	0.55
Fever/chills			0.85	0.58	0.70	0.83
Cough			0.90	0.94	0.80	0.49
Pleuritic chest pain			0.29	0.29	0.90	0.85
Dyspnea			0.44	0.78	0.87	0.45
Purulent sputum			0.65	0.42	0.95	0.82
X-ray						
Localized alveolar infiltrate			0.95	0.64	0.90	0.95

the sample of conditional statistics examined, there were both over- and underestimations. In the case of history, specificities were overestimated in the original frames. However, as table 3 indicates, the estimates of the sensitivities appeared essentially accurate. In the case of the radiographic data the sensitivities were significantly overestimated. The specificities, on the other hand, were generally underestimated.

To test the effects of revising the frame statistics, the original diagnostic system was run against the 100-member test set, then was revised using the statistics derived from the training group and was run again. The results are summarized in tables 5, 6, and 7.

Table 5 compares the accuracies of the differential lists for frames using the original and revised statistics. Three types of results are shown. First, the accuracies of the differential lists are compared for all 110 diagnoses in the 100 patients. This includes the diagnosis "no pulmonary disease." In frames using only historical information to estimate the likelihood of the diseases, the original frames captured 87 of the diagnoses while the revised frames included 94.

The second result displayed emphasizes system accuracy for the 58 pulmonary diseases found in this patient population. Of these diseases, 41 were included in the differential lists using the original frames. The revised frames captured 43 of the diseases.

The third approach to measuring accuracy looks at the disease ranked first rather than the five-member differential diagnostic list. The accuracy of this system using the original frames was 54 of 110 diseases. The revised frames yielded 60 of these diagnoses. This approach has the disadvantage of giving a low assessment of accuracy in populations where patients may have more than one disease. Only one of the diagnoses can be first.

None of the results was statistically significant at the 0.05 level. The change in the accuracy in the recognition of all diagnoses appeared to show a trend with a significance of < 0.10 using McNemar's test.

Table 6 shows the results of a similar evaluation of the test population using only data from the patient's first chest x-ray. The results reflect the magnitude of the change in the statistical underpinning of the ra-

diology portion of the revised frames. Comparison of the system for all diagnoses using the original frames yielded a success rate of 61 of 110 diseases; using the revised frames resulted in capture of 96 of the patients' diagnoses. The difference was significant at the 0.001 level using McNemar's test.

The change in system accuracy was also significant for the recognition of pulmonary diseases and for the evaluation of accuracy in terms of placing each of the 110 diagnoses first. The original system included ten of the patients' diseases in the differential lists, while the revised frames captured 44 of these 58 diseases ($p < 0.001$, McNemar's test).

In the third test, the original system succeeded in placing the patient's disease first in 45 instances while the system utilizing the revised frames placed it first in 62 cases. The difference was also significant ($p < 0.005$, McNemar's test).

In the final group of tests, both the history collected from the patient and the results of chest radiography were submitted to the original and revised systems. Both sets of frames did quite well, and no statistically significant difference was found. The results are shown in table 7.

Table 5 • Diagnoses Captured in Differential Lists Based on History Alone

	Original Frames	Revised Frames
All diagnoses (n = 110)	87 (79%)	94 (85%)*
Pulmonary diagnoses (n = 58)	41 (71%)	43 (74%)
Each diagnosis listed first (n = 110)	54 (49%)	60 (55%)

*$p < 0.10$ by McNemar's test.

Table 6 • Diagnoses Captured in Differential Lists Based on X-ray Alone

	Original Frames	Revised Frames
All diagnoses (n = 110)	61 (55%)	96 (87%)*
Pulmonary diagnoses (n = 58)	10 (17%)	44 (76%)*
Each diagnosis ranked first (n = 110)	45 (41%)	62 (56%)†

*$p < 0.001$ by McNemar's test.
†$p < 0.005$ by McNemar's test.

Table 7 • Diagnoses Captured in Differential Lists Based on History and X-ray Results

	Original Frames	Revised Frames
All diagnoses (n = 110)	96 (87%)	99 (90%)
Pulmonary diagnoses (n = 58)	46 (79%)	49 (84%)
Each diagnosis listed first (n = 110)	73 (66%)	70 (64%)

When the disease lists were examined for all diagnoses, the accuracy rate changed from 96 of 110 to 99 of 110. For the 58 pulmonary diagnoses, there were 46 successes produced by the original system and 49 produced by the revised system. For each diagnosis listed first, the original system captured 73, while the revised system found 70.

Discussion

In considering this set of diseases, the physicians consistently underestimated the a priori likelihood. Since we could only analyze the more common diseases, bias in this direction is unlikely to be consistent throughout the disease set. Although the selection criterion for our samples, hospitalized patients who had chest x-rays, does reflects a major part of the daily experience of the radiologists on our expert panel (three of the five physicians), it does not reflect the typical incidence/prevalence information in the medical literature, and this may account for the inaccuracy. Nonetheless, errors of this magnitude certainly effected system performance. Since a priori probabilities may vary substantially among groups of patients, analysis of sets of patients similar to the target group is the only reliable way to capture a priori probabilities.

Substantial estimation errors also appeared in the conditional probabilities. These errors are notably more pronounced in the probabilities for radiology than in those for history. This may reflect the fact that the estimates for history, particularly the sensitivities, were based in part on a group of statistics previously extracted from the HELP clinical database. This information was produced during the course of a previous test of computer-based history-collection tools.[3] Thus, the diagnostic results observed after revising the statistics for the history may not reflect the potential value of derived sensitivities and specificities as well as do the results associated with altering these values for x-ray findings.

We suspect that the estimation errors seen represent the general human difficulty with statistical estimation noted by Tversky and Kahneman.[7] The identification and analysis of a representative group of patients can avoid these errors.

The importance of errors of estimation in a Bayesian system is illustrated by the improvement seen in the behavior of our system for generating a differential diagnostic list when statistics derived from a database were substituted for estimated values. Improvement was evident in all of the subsets of data and approaches except the last analysis for the combined data. When only x-ray data were used the improvements were all statistically significant. Because the two sets of statistics for the history were quite similar, the differences between the original and revised frames

were not as striking as in the case of x-ray data or of the combined data from history and x-ray. However, analysis of larger data sets where only history was available confirms that the differences seen can in fact reach statistical significance.

This analysis emphasizes the usefulness of a clinical database in generating the values needed for a statistical approach to diagnostic decision making. While medical experts are reliable sources of information concerning which data are likely to be diagnostically useful, the sensitivities and specificities estimated by them produce less accurate systems than do those generated from a population similar to the one from which new patients are coming. This problem can be remedied by storing relevant data in a clinical database and using them to upgrade the diagnostic frames.

Our conclusion is that in order to use a statistical inference mechanism effectively in medical diagnosis, the diagnostic system must be closely tied to a clinical database representative of the types of patients and diseases on which the system is expected to operate. The ideal system would have some mechanism available automatically to review the clinical material and revise the diagnostic logic at fixed intervals.

References

1. Pozen MW, D'Agostino RB, Mitchell JB, et al. The usefulness of a predictive instrument to reduce inappropriate admissions of the coronary care unit. Ann Inter Med. 1980;92:238–42.
2. Faught E, Trader SD, Hanna GR. Cerebral complications of angiography for transient eschemia and stroke: prediction of risk. Neurology. 1979;29:4–15.
3. Warner HR, Rutherford BD, Houtchens BA. A sequential Bayesian approach to history taking and diagnosis. Comput Biomed Res. 1972;5:256–62.
4. DeDombal FT, Leaper DJ, Horrocks JC, et al. Human and computer-aided diagnosis of abdominal pain: further report with emphasis on performance of clinicians. Br Med J. 1974;1:376–80.
5. Hlatky M, Botvinick E, Brundage B. Diagnostic accuracy of cardiologists compared with probability calculations using Bayes' rule. Am J Cardiol. 1982;49:1927–31.
6. Osborne SF. Medical diagnosis aboard submarines—use of a computer-based Bayesian method of analysis in an abdominal pain diagnostic program. J Occup Med. 1984;26:110–4.
7. Tversky A, Kahneman D. Judgement under uncertainty: heuristics and biases. Science. 1974;185:1124–31.
8. Fineberg HV, Bauman R, Sosman M. Computerized cranial tomography—effect on diagnostic and therapeutic plans. JAMA. 1977;238:224–7.
9. Thornbury JR, Fryback DG, Edwards W. Likelihood ratios as a measure of the diagnostic usefulness of excretory urogram information. Radiology. 1975;114:561–5.
10. Leaper DJ, Horrocks JC, Staniland JR, DeDombal FT. Computer-assisted diagnosis of abdominal pain using "estimates" provided by clinicians. Br Med J. 1972;4:350–4.
11. Greibe J, Bugge P, Gjorup T, Lauritzen T, Bonnevie O, Wulff HR. Long-term prognosis of duodenal ulcer: follow-up study and survey of doctors' estimates. Br Med J. 1977;2:1572–4.
12. Ben-Bassat M, Carlson RW, Puri VK, et al. Pattern-based interactive diagnosis of multiple disorders: the MEDAS system. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-2, No. 2, March 1980.
13. International classification of disease. 9th revision, clinical modification, second edition. U.S. Department of Health and Human Services, 1980.

Appendix

Variations on Bayes equation have been extensively used for developing statistical models of the diagnostic process. The version of Bayes equation typically used in medical diagnosis is:

$$P(D|F) = \frac{P(D)P(F|D)}{P(F)} = \frac{P(D)P(F|D)}{\sum_i P(D_i)P(F|D_i)} \quad (1)$$

where $P(D|F)$ represents the probability of diagnosis D when finding F is present, $P(D)$ is the probability of this disease prior to the recognition of finding F, $P(F|D)$ is the frequency of occurrence of finding F in patients with disease D (*sensitivity*), and $P(F)$ is the frequency of finding F in the entire population under consideration. As indicated $P(F)$ is typically expanded to a sum made up of the products of the probabilities of each of the possible diseases in this population ($P(D_i)$) and the individual sensitivities of the findings in each of these diseases.

The multi-membership version of Bayes equation uses a different formulation for $P(F)$:

$$P(D|F) = \frac{P((D)P(F|D)}{P(D)P(F|D) + P(\overline{D})P(F|\overline{D})} \quad (2)$$

In this case the denominator has been replaced by the sum of two products, the product of the probability of the disease and the sensitivity, and the product of the probability of being free of the disease, $P(\overline{D}) = 1 - P(D)$ and the false-positive rate, $P(F|\overline{D})$.

The principal advantage of this version of Bayes equation is that, unlike equation (1) it does not require the assumption that the diseases considered by the system are mutually exclusive and exhaustive. In our experience, patients frequently have more than one of the diseases included in this system. The disadvantage of using equation (2) in a diagnostic system is the necessity of capturing an additional statistic from the medical experts. A false-positive rate (which is equal to $1 - P(\overline{F}|\overline{D})$ or $1 -$ *specificity*) is required for each symptom used in a disease module. In practice the specificity tends to be a more familiar number and is frequently easier to collect from the experts who populate the knowledge base.